Int J Life Cycle Assess (2014) 19:1445–1461
DOI 10.1007/s11367-014-0751-0

UNCERTAINTIES IN LCA

Error propagation methods for LCA—a comparison

Reinout Heijungs · Manfred Lenzen

Received: 22 March 2013 / Accepted: 6 May 2014 / Published online: 28 May 2014

Abstract
Purpose The analysis of uncertainty in life cycle assessment (LCA) studies has been a topic for more than 10 years, and many commercial LCA programs now feature a sampling approach called Monte Carlo analysis. Yet, a full Monte Carlo analysis of a large LCA system, for instance containing the 4,000 unit processes of ecoinvent v2.2, is rarely carried out by LCA practitioners. One reason for this is computation time. An alternative faster than Monte Carlo method is analytical error propagation by means of a Taylor series expansion; however, this approach suffers from being explained in the literature in conflicting ways, hampering implementation in most software packages for LCA. The purpose of this paper is to compare the two different approaches from a theoretical and practical perspective.
Methods In this paper, we compare the analytical and sampling approaches in terms of their theoretical background and their mathematical formulation. Using three case studies—one stylized, one real-sized, and one input–output (IO)-based—we approach these techniques from a practical perspective and compare them in terms of speed and results.
Results Depending on the precise question, a sampling or an analytical approach provides more useful information. Whenever they provide the same indicators, an analytical approach is much faster but less reliable when the uncertainties are large.
Conclusions For a good analysis, analytical and sampling approaches are equally important, and we recommend practitioners to use both whenever available, and we recommend software suppliers to implement both.

Keywords Analytical methods · Gaussian error propagation · IOA · LCA · Monte Carlo · Sampling methods · Uncertainty

1 Introduction

Uncertainty is a pervasive topic in life cycle assessment (LCA). It is so in a fundamental sense: uncertainty is present in many forms in all stages of an LCA. It is also well presented in the recent scientific literature: the last few volumes of the *International Journal of Life Cycle Assessment* contain papers on uncertainty in LCA, either in recognizing that there is uncertainty or in presenting approaches to manage them. Review articles on uncertainty have been published by various authors (Heijungs and Huijbregts 2004; Lloyd and Ries 2007). One of the most used life cycle inventory (LCI) databases—ecoinvent—contains specifications of parameter uncertainty for many coefficients of most unit processes. Most software for LCA is by now able to deal with uncertainties, in most cases on the basis of Monte Carlo simulations. A UNEP-SETAC-endorsed working group on uncertainties in LCA has been active for a few years.

Responsible editor: Andreas Ciroth

R. Heijungs (✉)
Department of Industrial Ecology, Institute of Environmental Sciences, Leiden University, P.O. Box 9518, 2300 RA Leiden, The Netherlands
e-mail: heijungs@cml.leidenuniv.nl

R. Heijungs
e-mail: r.heijungs@vu.nl

R. Heijungs
Department of Econometrics and Operations Research, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

M. Lenzen
ISA, School of Physics A28, The University of Sydney, Sydney, NSW 2006, Australia

Uncertainty shows up in LCA in many ways:

- Input data (such as fuel consumption, CO_2 emissions, and characterisation factors) may be uncertain or conflicting due to inaccurate measurements, or they may be subject to variability from day to day or from source to source.
- The LCA procedure requires a number of choices and assumptions (for instance, on system boundaries, consequential vs. attributional LCA, and the time horizon for global warming), and such choices are debatable.
- Uncertain and variable data and different choices percolate through the LCA model via error propagation, leading to uncertain results.
- Uncertain results may be interpreted by decision-makers in varying ways, depending on preferences, time, or framing of the decision situation, amongst other factors.

In any case, uncertainty needs to be documented in LCA studies, and the framework shown in Fig. 1 can be used for this purpose.

This article addresses one specific issue pertaining to uncertainty in relation to LCA: the estimation of the uncertainty of LCA results using given uncertainty estimates of the LCA input parameters. This issue is represented by the step indicated by bold lines in Fig. 1. The subject field behind this is known as the theory of error propagation, as it addresses the question how input uncertainties propagate into output uncertainties through the LCA model. In this paper, we will restrict our attention to LCA based on unit process data and LCA based on input–output (IO) data. Customized, parameterized LCA models, in which for instance the fuel input and the emissions depend on a fuel efficiency parameter at a higher level, are not discussed in this paper. The chief purpose of this paper is to demonstrate to practitioners the principles, assumptions, and comparative advantages of the main methods for propagating parameter uncertainty to overall LCI uncertainty, under well-defined assumptions (i.e., no interaction between variables and relatively small and well-behaving uncertainties).

Basically, two classes of techniques are available to study error propagation: sampling methods and analytical methods.

Sampling methods are methods that address the problem by sampling from the probability distributions of the input parameters and re-calculating the LCA results for every member of this sample (Morgan and Henrion 1990). This produces a sample of results, and from this sample, several statistics can be computed, such as the mean, the standard deviation, the median, 95 % confidence intervals, correlation coefficients, etc. The most well-known sampling method is Monte Carlo simulation. A more sophisticated method is the Latin hypercube method, where the sampling strategy is not entirely random but utilizes stratified probability distributions. Yet, another approach features a calculation run for every combination of parameters in which each parameter assumes the highest and the lowest value.

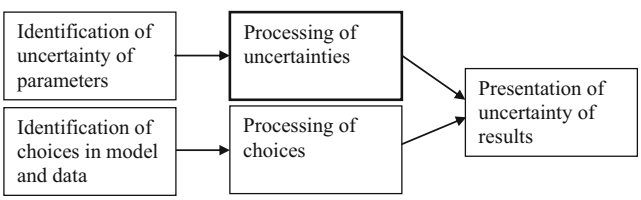

Fig. 1 Framework for treating parameter and model uncertainty in LCA. The *bold box* represents the emphasis of this paper

Sampling methods are conceptually easy to understand, and they are easy to implement in software. There exist commercial packages (such as Crystal Ball and @RISK) that can be used as a plug-in together with other software (such as Microsoft Excel). A brief survey suggests that many present-day LCA programs have implemented the Monte Carlo method, which naturally means that most LCA studies that include an uncertainty analysis do this on the basis of Monte Carlo simulation (Lloyd and Ries 2007).

Sampling methods, however, have also disadvantages. The number of simulations must be large enough to ensure sampling means are measured with small enough standard deviations. As a consequence, many Monte Carlo studies typically take 10,000 calculation runs. Using a Latin hypercube strategy may reduce this, but LCA programs do, as far as we know, not include this more advanced sampling method. For the conceptually simpler strategy of using all combinations of highest and lowest value, Heijungs (1996) has shown that the computation time required in LCAs may easily exceed the age of the universe. In a small LCA, the computation time for LCI and/or LCIA can typically be less than a second. Running 1,000 Monte Carlo simulations would then require a quarter of an hour. This is doable, although repeating this for many pollutants and/or many product alternatives is not an attractive option. Running 10,000 simulations would require 3 h, and that already is more of a problem. Present-day LCA inventories tend to be large. As an illustration, the 1996 version of the ETH energy database contains 1,200 processes, ecoinvent v1 from 2002 contains 2,500 processes, ecoinvent v2 (2007) 4,000 processes, and the recently released version v3 about 10,000 processes. One popular way of solving an LCI is the inversion of an $N \times N$ square matrix, and that this is typically an operation requiring N^3 single computational steps (Press et al. 1992). As a result, doubling a database in size implies an eightfold increase in computation time, all other things remaining equal. Indeed, whereas ecoinvent v1 was shipped with the results of Monte Carlo calculations, no such calculations have been performed for some of the versions of ecoinvent v2. A smarter sampling strategy, for example using a Latin hypercube approach, may reduce the required number of simulations needed. And a smarter algorithm for computing an LCI (Peters 2007) may reduce the required effort as well. But yet, even when one run

takes 1 min, unacceptably large computation times are required for day-to-day consultants' purposes.

An alternative approach to study the propagation of uncertainties is by using an analytical approach. Such an approach is more difficult to understand and requires more mathematical background, although once implemented into software, no mathematical expertise is required from the LCA practitioner. The analytical approach is based on calculus, applying a local derivative of the mathematical function that specifies how inputs are transformed into outputs. In contrast to a sampling method, it does not provide a probability distribution of the outputs. Rather, it only estimates the mean and the standard deviation (or its square, the variance). This, however, is sufficient for many purposes.

An initial study (Heijungs et al. 2005) compared the performance of the sampling and analytical approaches and concluded the following: "When one Monte Carlo run takes 30 s, 1,000 runs require a working day. The analytical approach can reduce this to a few minutes, while the results are basically the same" (p.111). This was, however, on one test system, using outdated software, and it was certainly not an exhaustive analysis of sampling vis-à-vis analytical methods.

The present paper approaches the two methods in more detail. We discuss three different ways of elaborating the analytical approach: one as published by Heijungs (1994), one by Hong et al. (2010), and one on the basis of input–output analysis, while the section thereafter discusses the sampling method in more detail. Section 4 presents three case studies in which the analytical method is compared to the sampling approach. Section 5 concludes the paper.

2 Analytical approaches to error propagation in LCA

In this section, we first discuss the theory of analytical error propagation (Section 2.1) and then introduce a number of alternative formulations for error propagation (Section 2.2), notably the analytical method. Finally, we present the analytical expressions for the LCA model, according to different publications (Section 2.3).

2.1 Theory of error propagation

2.1.1 Taylor series expansion

Taylor's theorem is an established part of mathematical analysis (Apostol 1967). It asserts that we can calculate the value of a function f for a value x if we know it at another point a, using a number of terms that involves the derivatives of f with respect to x and using the distance between x and a. The approximation of $f(x)$ can be made more precise by taking into account more of these terms, viz., by taking a longer expansion. The precise form is as follows:

$$f(x) = f(a) + \frac{df(a)}{dx}(x-a) + \frac{d^2f(a)}{dx^2}(x-a)^2 + \frac{d^3f(a)}{dx^3}(x-a)^3 + \cdots \quad (1)$$

This approximation does not hold for every function f, but it requires some assumptions (such as differentiability) that we will assume to hold in the cases we discuss hereafter. An example is the Taylor series expansion for $\sin(x)$ around the value $a=0$: $\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots$. For small values of x, we can thus use $\sin(x)=x$, for somewhat larger values of x; we need to take account of the second term as well. In this way, we can use within a certain region a simple linear approximation for any complicated function. When we use the first term only, we speak of a first-order approximation, including the second term gives a second-order approximation, etc.

2.1.2 Determining variance—an illustrative example

To explain the analytical error propagation method, we first take a simple example. Suppose we want to calculate the area A of a rectangular sheet of paper with width w and height h. The formula to do this is simply

$$A = w \times h \quad (2)$$

Let us suppose that both w and h are empirically determined variables, which are not exactly known, but which are known to be specified according to probability distributions with variances var(w) and var(h), where var is the square of the standard deviation of the probability distribution, often written as s^2 or σ^2. Further, we assume that the errors in w and h are independent, i.e., that the covariance of w and h is 0 (we discuss the plausibility and consequences of this assumption in the last section). It is now asked to specify the probability distribution of A and in particular its variance var(A). The theory of error propagation (Bevington and Robinson 1994; Morgan and Henrion 1990; Ayyub and Klir 2006) gives an answer to this:

$$\mathrm{var}(A) \approx h^2 \times \mathrm{var}(w) + w^2 \times \mathrm{var}(h), \quad (3)$$

where the $\approx$ indicates that it is an approximate value, as we have only used a first-order approximation and have neglected any possible correlation between the values of w and h. Thus, given the mean values of w and h and the variances of w and h, the mean and the variance of A can be computed using two simple formulas.

This is a special and simple case. The more general form in the case of two input variables x and y is that of an output

variable z that is an arbitrary function of the two input variables,

$$z = z(x, y).$$

In that case, we have

$$\mathrm{var}(z) \approx \left(\frac{\partial z}{\partial x}\right)^2 \mathrm{var}(x) + \left(\frac{\partial z}{\partial y}\right)^2 \mathrm{var}(y) + 2\left(\frac{\partial z}{\partial x}\right)\left(\frac{\partial z}{\partial y}\right) \mathrm{cov}(x, y), \quad (4)$$

where we have now included the possibility that the errors of the variables x and y are correlated, expressed through the covariance (cov) of x and y (see, e.g., Bevington and Robinson 1994). It is now a task to find the expression for the function $z(x, y)$ in LCA. This will be discussed in the next section.

2.1.3 Assumptions and definitions

To study the analytical approach toward error propagation in LCA, we will first make several assumptions first.

- We focus on the LCI phase. Extension of these ideas to cover characterization, normalization, weighting, etc. are straightforward, albeit they require hard work (Ciroth et al. 2004).
- We restrict the discussion to parameter uncertainty. Model uncertainty and other forms of uncertainty that may be involved in setting system boundaries, allocation, etc. are excluded here.
- We consider only random errors, not systematic errors. If known, systematic errors should be corrected before an uncertainty analysis, which is easily possible, also for LCAs (Bevington (1994), p. 41; Ciroth (2001, pp. 126).

In order to discuss LCA in mathematical terms, some conventions and symbols must be chosen. Following the symbols introduced by Heijungs and Suh (2002), we use the following:

- Bold lowercase letters (like **z**) for vectors, bold capital letters (like **Z**) for matrices, and italic letters (like z) for scalars
- The product (and material, energy, …) requirement and supply of unit processes are specified as a column vector **a**, in which an element a_i indicates the use or supply of product i
- A positive coefficient a_i means production, a negative coefficient means use, a zero value means that a product is not involved in a process
- The environmental exchanges (the elementary flows) of unit processes are specified as a column vector **b**, in which an element b_l indicates the use of resource l or the emission of pollutant l
- Different unit processes can be combined in an LCA system by horizontally concatenating column vectors into matrices **A** and **B**
- The functional unit is defined in the form of a vector **f**, in which the reference flow (for product i) is represented as the only non-zero element f_i
- The inventory result is defined as a vector **g**, of which an element g_l indicates the system-wide use of resource l or the system-wide emission of pollutant l

2.2 Review of alternative error propagation formulations

Even though the form of the function **g**(**B**, **A**, **f**) is an elementary question in the calculation of LCA, it has remarkably little been addressed (Heijungs and Suh 2002). Most LCI and LCA textbooks do not discuss it or only superficially mention the issue without proposing a specific formula.

Several solutions to this question have been provided by different authors.

- Heijungs et al. (1992, 1994) use Cramer's rule.
- Heijungs (1996) and Heijungs and Suh (2002) use matrix inversion.
- Ciroth (2001, 2004) uses a sequential calculation approach. Due to complications with looped systems (which we believe is the default situation), this will not be discussed in this paper.
- Hong et al. (2010) and Imbeault-Tétrault et al. (2013) do not discuss the form of **g**, but they do discuss the form of its derivative in the context of uncertainty propagation.

These differences in approach naturally lead to differences in the form of the equation for **g** and hence in the equations for its derivatives with respect to **B**, **A**, and **f**, i.e., $\frac{dg_k}{db_{lj}}$, $\frac{dg_k}{da_{ij}}$, and $\frac{dg_k}{df_i}$.

2.2.1 LCI according to Heijungs

In a series of contributions, Heijungs et al. have developed explicit equations for the determination of **g** and its derivatives.

Already in the Dutch LCA guidebook (Heijungs et al. 1992), the question of calculating a life cycle inventory is presented as one of solving a system of linear equations. Under the usual assumptions of linear technology, market clearing (product balance), steady state, an equal number of processes and products, etc., the model equations can be written as follows:

$$\begin{cases} \mathbf{f} = \mathbf{As} \\ \mathbf{g} = \mathbf{Bs} \end{cases} \quad (5)$$

where the bracket indicates that these two equations hold simultaneously. The first system of equations specifies the known **f** and **A**, and the challenge is to solve for the elements

of **s**, which are interpreted as the intensities, levels, or scaling factors of the processes in the system. The result can be fed into the second equation, to find **g** by straightforward multiplication.

Heijungs et al. (1992) and Heijungs (1994) adopt Cramer's rule to solve the first equation. In later publications, Heijungs et al. turned from Cramer's rule to matrix inversion (Heijungs 1996; Heijungs and Suh 2002). In these publications, the equations are

$$\mathbf{s} = \mathbf{A}^{-1}\mathbf{f} \tag{6}$$

and

$$\mathbf{g} = \mathbf{B}\mathbf{A}^{-1}\mathbf{f} = \Lambda\mathbf{f}. \tag{7}$$

Heijungs and Suh (2002, p. 133 ff.) provide equations for the derivatives. For instance, it is found that

$$\frac{\partial g_k}{\partial a_{ij}} = \left(\mathbf{B}\mathbf{A}^{-1}\right)_{ki}\left(\mathbf{A}^{-1}f\right)_j = \lambda_{ki}s_j \tag{8}$$

A similar formula is discussed for $\frac{\partial g_k}{\partial b_{ij}}$. The one for $\frac{\partial g_k}{\partial f_i}$ is assumed to be zero, as the final demand vector is user-defined and free from error in most LCA studies. Inserted in the formula for error propagation [4] without covariance term, we easily find

$$\mathrm{var}(g_k) = \sum\nolimits_{i,j}\left(s_j\lambda_{ki}\right)^2\mathrm{var}\left(a_{ij}\right) + \sum\nolimits_j\left(s_j\right)^2\mathrm{var}\left(b_{kj}\right) \tag{9}$$

Heijungs (2010) provides a systematic overview of all derivatives that can show up in LCA.

2.2.2 Input–output-based LCI

During the past 10 years, economic IO techniques have made their way into LCA as an essential ingredient of hybrid methods combining IO-based and process-based methods (Suh and Nakamura 2007). The perceived advantage of this hybridization is that the best of both worlds—infinite upstream coverage of IO and detail and specificity of process methods—are combined in one approach (Bullard et al. 1978; Moskowitz and Rowe 1985; Suh et al. 2004). Hybrid LCI methods have been described in detail by Heijungs and Suh (2002), Suh (2004), and Suh and Huppes (2005). Williams et al. (2009) propose a hybrid approach combining process and IO approaches to LCI uncertainty analysis.

IO data are, like process-based data, subject to uncertainty. Below, we will elaborate the IO setup for an industry-by-industry monetary table in coefficient form, with an emphasis on the connection with uncertainty propagation. Other setups (product-by-product, physical, transaction form) are easy to derive from the base case.

Let one element of the environmental repercussions **b** (e.g., sectoral CO_2 emission) of an arbitrary functional unit vector **f** (e.g., a household's final consumption) defined as $g=\mathbf{b}^t\mathbf{L}\mathbf{f}=\mathbf{b}^t(\mathbf{I}-\mathbf{A})^{-1}\mathbf{f}$, where $\mathbf{L}=(\mathbf{I}-\mathbf{A})^{-1}$ is the Leontief inverse of the direct requirements or technical coefficient matrix **A**. The matrix **A** here contains only input data (as positive numbers), while **A** in the process-based setup represents both inputs (negative) and outputs (positive). The implicit output of every sector is 1; hence, the "**I**" and the sign reversal explains the "–" in front of the **A**.

Ignoring variations in **b** and **f**, the Taylor series of g as a function $g(\mathbf{A})$ is

$$\begin{aligned} g(\mathbf{A}) \approx g(\mathbf{A}_0) &+ \sum_{ij}\left(A_{ij}-A_{0,ij}\right)\frac{\partial g}{\partial A_{ij}}(\mathbf{A}_0) \\ &+ \frac{1}{2!}\sum_{ij}\sum_{k,l}\left(A_{ij}-A_{0,ij}\right)\left(A_{kl}-A_{0,kl}\right)\frac{\partial^2 g}{\partial A_{ij}\partial A_{kl}}(\mathbf{A}_0) \\ &= g(\mathbf{A}_0) + \mathbf{1}^{\mathrm{t}}(\mathbf{A}-\mathbf{A}_0)\#\mathbf{D}g(\mathbf{A}_0)\mathbf{1} + \frac{1}{2!}\mathbf{1}^{\mathrm{t}}(\mathbf{A}-\mathbf{A}_0)\#\mathbf{D}^2g(\mathbf{A}_0)\#(\mathbf{A}-\mathbf{A}_0)\mathbf{1}, \end{aligned} \tag{10}$$

where # denotes the element-wise product, and $\mathbf{1}^t=(1,1,\ldots,1)$ is a suitable summation operator. $\mathbf{D}g(\mathbf{A}_0)$ is called the gradient of g at $\mathbf{A}_0$, and $\mathbf{D}^2g(\mathbf{A}_0)$ is called the Hessian of g at $\mathbf{A}_0$. For $g(\mathbf{A})$ as above, we find for example

$$\begin{aligned} \frac{\partial g}{\partial A_{ij}} &= \frac{\partial\left(\mathbf{b}^t\mathbf{f} + \mathbf{b}^t\mathbf{A}\mathbf{f} + \mathbf{b}^t\mathbf{A}^2\mathbf{f} + \mathbf{b}^t\mathbf{A}^3\mathbf{f} + \ldots\right)}{\partial A_{ij}} \\ &= 0 + b_if_j + \left[\left(\mathbf{b}^t\mathbf{A}\right)_if_j + b_i\left(\mathbf{A}\mathbf{f}\right)_j\right] + \left[\left(\mathbf{b}^t\mathbf{A}^2\right)_if_j + \left(\mathbf{b}^t\mathbf{A}\right)_i\left(\mathbf{A}\mathbf{f}\right)_j + b_i\left(\mathbf{A}^2\mathbf{f}\right)_j\right] \\ &\quad + \left[\left(\mathbf{b}^t\mathbf{A}^3\right)_if_j + \left(\mathbf{b}^t\mathbf{A}^2\right)_i\left(\mathbf{A}\mathbf{f}\right)_j + \left(\mathbf{b}^t\mathbf{A}\right)_i\left(\mathbf{A}^2\mathbf{f}\right)_j + b_i\left(\mathbf{A}^3\mathbf{f}\right)_j\right] + \ldots \\ &= \left[b_i + \left(\mathbf{b}^t\mathbf{A}\right)_i + \left(\mathbf{b}^t\mathbf{A}^2\right)_i + \ldots\right]\left[f_j + \left(\mathbf{A}\mathbf{f}\right)_j + \left(\mathbf{A}^2\mathbf{f}\right)_j + \ldots\right] = \left(\mathbf{b}^t\mathbf{L}\right)_i\left(\mathbf{L}\mathbf{f}\right)_j \end{aligned} \tag{11}$$

and

$$\{\mathbf{Dg}(\mathbf{A}_0)\}_{ij} = \left[\mathbf{b}^{\mathrm{t}}(\mathbf{I}-\mathbf{A}_0)^{-1}\right]_i \left[(\mathbf{I}-\mathbf{A}_0)^{-1}\mathbf{f}\right]_{\mathrm{j}}, \tag{12}$$

so that up to first order

$$g(\mathbf{A})-g(\mathbf{A}_0)\approx\sum_{ij}\left(A_{ij}-A_{0,ij}\right)\left[\mathbf{b}^t(\mathbf{I}-\mathbf{A}_0)^{-1}\right]_i\left[(\mathbf{I}-\mathbf{A}_0)^{-1}\mathbf{f}\right]_j \tag{13}$$

g can also be understood as a function of the input–output intermediate transactions matrix $\mathbf{T}=\mathbf{Ax}$ (where $\mathbf{x}$ is total use by sector, see Section 4.3), and a similar approach can be taken for this option as well. With

$$\frac{\partial g}{\partial A_{ij}} = \frac{\partial g}{\partial T_{ij}}\frac{\partial T_{ij}}{\partial A_{ij}} \Leftrightarrow \frac{\partial g}{\partial T_{ij}} = \frac{(\mathbf{b}^t\mathbf{L})_i(\mathbf{Lf})_j}{x_j} \tag{14}$$

we find the first-order approximation of $g(\mathbf{T})$

$$g(\mathbf{T})-g(\mathbf{T}_0)\approx\sum_{ij}\left(T_{ij}-T_{0,ij}\right)\frac{\left[\mathbf{b}^t\left(\mathbf{I}-\mathbf{T}_0\widehat{\mathbf{x}}^{-1}\right)^{-1}\right]_i\left[\left(\mathbf{I}-\mathbf{T}_0\widehat{\mathbf{x}}^{-1}\right)^{-1}\mathbf{f}\right]_j}{x_j} \tag{15}$$

Similar relationships can be derived for **b** and **f**. Note that **b** and **f** can also be multi-row and multi-column matrices, handling error propagation simultaneously for multiple interventions and functional units.

2.2.3 LCI according to Hong

A completely different approach is the one taken by Hong et al. (2010). These authors do not start from the equation for LCI to calculate the derivatives and insert them into the Taylor formula. Rather, they essentially skip the LCI equation and they even do not calculate explicit derivatives. Their approach is based on the idea that most data in LCA follow a log-normal distribution. Ignoring the covariance, the first-order Taylor series approximation is

$$\mathrm{var}(z)\approx\left(\frac{\partial z}{\partial x}\right)^2\mathrm{var}(x)+\left(\frac{\partial z}{\partial y}\right)^2\mathrm{var}(y) \tag{16}$$

By log-transforming all variables

$$x=\ln x';\ y=\ln y';\ z=\ln z' \tag{17}$$

this yields

$$\mathrm{var}(\ln z')\approx\left(\frac{\partial \ln z'}{\partial \ln x'}\right)^2\mathrm{var}(\ln x')+\left(\frac{\partial \ln z'}{\partial \ln y'}\right)^2\mathrm{var}(\ln y') \tag{18}$$

The terms with derivatives in Eq. [18] can be worked out with the derivative of the ln:

$$\frac{\partial \ln z'}{\partial \ln x'} = \frac{x'\,\partial z'}{z'\,\partial x'} = S_{z,x} \tag{19}$$

Further, the standard deviation of a stochastic variable x, SD(x), is related to its geometric standard deviation, GSD(e^x), by

$$\mathrm{SD}(x) = \ln(\mathrm{GSD}(e^x)) = \ln(\mathrm{GSD}(x')). \tag{20}$$

Inserting this for SD(x) and SD(y) yields

$$\ln(\mathrm{GSD}(z'))^2 = S_{z,x}\ln(\mathrm{GSD}(x'))^2 + S_{z,y}\ln(\mathrm{GSD}(y'))^2. \tag{21}$$

This is still a general form. For the case of LCA, the more specific expression

$$\ln(\mathrm{GSD}(g_k))^2 = \sum\nolimits_{i,j} S_{k,i,j}\ln\left(\mathrm{GSD}\left(a_{ij}\right)\right)^2 + \sum\nolimits_j \mathrm{S}_{k,j}\ln\left(\mathrm{GSD}\left(b_{kj}\right)\right)^2 \tag{22}$$

appears. The formula still contains two relative sensitivity coefficients, defined by

$$S_{k,i,j} = \frac{\partial g_k}{\partial a_{ij}}\frac{a_{ij}}{g_k} \text{ and } S_{k,j} = \frac{\partial g_k}{\partial b_{kj}}\frac{b_{kj}}{g_k}. \tag{23}$$

These coefficients can be calculated using the analytical results of Heijungs (2010) or by repeated calculation, approximating, for instance

$$S_{k,i,j} = \frac{\Delta g_k\, a_{ij}}{\Delta a_{ij}\, g_k} \tag{24}$$

with Δa_{ij}=1 %.

According to Hong et al. (2010), the formulas apply "to the case where both input and output are log-normally distributed." It is, however, unclear where this assumption is needed in the derivation above. A similar approach is taken by Imbeault-Tétrault et al. (2013), who list and discuss the assumptions: a multiplicative model, log-normally distributed input parameters, and independently distributed input parameters. We agree with these authors that the model is not multiplicative but with an even further-going argument: a matrix inverse resembles a division more than a multiplication. But in contrast to Imbeault-Tétrault et al. (2013), we do not see the need for making the assumption of log-normality because the formula for Gaussian error propagation nowhere makes restrictions on the shape of the distribution. Finally, we do agree with their

assumption of independence of input distributions, not because it is true (it is not) but because the uncertainty model would become too complicated and the data demand would become too large. Because we disagreed with the assumption of log-normality, we have not further pursued this approach any further. In a follow-up, a comparison between different analytical approaches may be carried out.

2.3 Some key metrics used in analytical error propagation

2.3.1 Non-comparative metrics

The approach described in Section 2.2 can be followed for every environmental flow, for every alternative. For each of these, it does not return a distribution. It only delivers a variance var(z) of some variable of interest as basic material. This variance can be used to calculate the standard deviation

$$\mathrm{sd}(z) = \sqrt{\mathrm{var}(z)} \tag{25}$$

and the coefficient of variation

$$\mathrm{CV}(z) = \frac{\mathrm{sd}(z)}{\mathrm{m}(z)}, \tag{26}$$

where we use the (only) value of z for the expectation value of z:

$$m(z) = z. \tag{27}$$

Because we do not know anything of the distribution of the series of results z, it is impossible to calculate a 95 % confidence interval, a significance level, etc. That is different for Hong et al. (2010), who assume that the output distribution in LCA is log-normal. By introducing this assumption and calculating a measure of dispersion (in their case, a geometric standard deviation), they can calculate a confidence interval and test the null hypothesis that the true value is zero.

2.3.2 Comparative metrics

Many LCA studies are comparative, that is, they aim to compare the environmental performance of two products I and II or even of a larger set of products I, II, III, etc. (Lenzen 2006). As discussed by several authors (see, e.g., Hong et al. 2010), there is an issue in the case of comparative LCA. When we wish to compare two alternatives, I and II, we are interested in the value of a certain result, say z, for both systems. In a traditional LCA, without uncertainties, we calculate z_I and z_II and see which one is higher. When we calculate the uncertainties with the analytical method, we find in addition a measure of the dispersion of the central values for I and II. But we cannot test whether the difference is in any sense significant or not. Again, the assumption of log-normality by Hong et al. (2010) and Imbeault-Tétrault et al. (2013) opens new vistas. These authors build on the comparison indicator by Huijbregts (1998), defined as

$$\mathrm{CI}_{\mathrm{I,II}}(z) = \frac{z_\mathrm{I}}{z_\mathrm{II}}. \tag{28}$$

As z_I and z_II are assumed to follow a log-normal distribution, CI follows this distribution as well. With the additional assumption that z_I and z_II are independent, it follows that

$$\ln^2\left(\mathrm{GSD}\left(\mathrm{CI}_{\mathrm{I,II}}(z)\right)\right) = \ln^2(\mathrm{GSD}(z_\mathrm{I})) + \ln^2(\mathrm{GSD}(z_\mathrm{II})) \tag{29}$$

so that a measure of dispersion (the GSD) of the comparison indicator can be obtained. Unfortunately, when z_I and z_II are conceived as independent variables, the ratio $\mathrm{CI}_{\mathrm{I,II}}$ has a higher uncertainty than those of the contributing items, whereas our conjecture is that the comparison will have a lower uncertainty in practice, due to the fact that two scenarios I and II share a background system with common uncertainties (cf. Eq. (6) in Imbeault-Tétrault et al. 2013).

2.3.3 Contribution to variance

The uncertainty of a result can be conceived as being built up by uncertainties of the parameters. Some of these parameters are fairly certain, but even when all parameters are equally uncertain, some of them will have a higher influence on the uncertainty of the output than others. An analysis of the contribution to variance (CTV) decomposes the uncertainty in its contributing uncertainties. This provides important information for efficiently reducing the uncertainty.

Given the basic formula [4] neglecting covariance

$$\mathrm{var}(z) \approx \left(\frac{\partial z}{\partial x}\right)^2 \mathrm{var}(x) + \left(\frac{\partial z}{\partial y}\right)^2 \mathrm{var}(y) \tag{30}$$

we can easily express the contribution to variance of z by the uncertainty in x, CTV(z,x), as

$$\mathrm{CTV}(z,x) = \left(\frac{\partial z}{\partial x}\right)^2 \mathrm{var}(x), \tag{31}$$

and similarly for CTV(z,y). For the case of LCA, following Heijungs' approach, ones finds

$$\mathrm{CTV}\left(g_k, a_{ij}\right) = \left(s_j \lambda_{ki}\right)^2 \mathrm{var}\left(a_{ij}\right) \tag{32}$$

and

$$\mathrm{CTV}\left(g_k, b_{kj}\right) = \left(s_j\right)^2 \mathrm{var}\left(b_{kj}\right) \tag{33}$$

Typically,

$$\mathrm{CTV}(g_k, f_i) = 0 \tag{34}$$

because the functional unit is a fixed number.

3 Sampling methods for error analysis

Sampling methods are based on repeatedly running the model with a sample of input parameters, to create a sample of model results. This sample of model results can be analyzed to extract uncertainty indicators. This section introduces different aspects of the sampling method: the main idea of Monte Carlo sampling (Section 3.1), some metrics that can be derived from a sample of results (Section 3.2 for the non-comparative case and Section 3.3 for the comparative case), some details on the implementation of Monte Carlo sampling in LCA (Section 3.4), and the possibilities for extracting contribution to variance information from a sample of results (Section 3.5).

3.1 Monte Carlo analysis

Suppose that we have specified all input parameters as distributions, e.g., the coefficients w and h in the example on the sheet of paper (Section 2.1.2) are not fixed numbers but are defined in terms of probability distributions with specified type and parameters. In one run, i, we draw all parameters from the specified distribution, defined by $(w,h)_i$.

We calculate quantities of interest, in this case $A_i = w_i \times h_i$ (see Eq. [2]). Repeating this N times, we obtain a sample $\{A_1, A_2, \cdots, A_N\}$. This is the basic idea of Monte Carlo sampling. The method assumes that the probability distribution is sampled in a representative way. This may require a substantial number of runs.

In the case of LCA, the procedure is similar. Here, the input data consists of the matrices **A** and **B** and the output of the vector **g** (see, for instance, Eq. [7]). The Monte Carlo sample obtained is thus $\{\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_N\}$.

More sophisticated methods exist to explore the sampling space in a more informed way. Using structural path analysis (Crama et al. 1984; Defourny and Thorbecke 1984; Lenzen 2002; Suh and Heijungs 2007) or the path exchange method (Treloar 1997; Lenzen and Crawford 2009), an LCI system can be reduced to its most significant pathways and nodes, resulting in a much simpler and more rapid sampling task. We have not explored these other techniques in this paper.

Monte Carlo techniques have been used in economic input–output analysis for at least 50 years (Quandt 1958; 1959) and in environmental input–output analysis for at least 30 years, recently including hybrid LCA (Bullard and Sebald 1977, 1988; Ii 2000; Sakai et al. 2000; Nansai et al. 2001; Yoshida et al. 2001; Yoshida et al. 2002). These techniques can be applied to a number of error categories, such as source data uncertainty and aggregation error, amongst others (Lenzen 2000).

In carrying out a Monte Carlo analysis and in analyzing its results, we should distinguish two cases:

- The one-sample or independent sample case, where a result is calculated for one product, or for several products but using new Monte Carlo runs
- The multiple-sample case, where one Monte Carlo realization is used to calculate an observation for several products

In discussing the analytical approach above, we treated the comparative and non-comparative metrics separately. In discussing the sampling approach, we will do this again, but here, the options are richer because the samples may originate from independent or dependent stochastic runs.

3.2 Non-comparative metrics calculated with the sampling approach

A single sample of results may be analyzed by calculating various statistics. We subdivide these statistics into the family of parametric statistics and non-parametric statistics (Siegel 1956). Table 1 shows some key statistics that can be derived from a sample of results for one alternative. Appendix 1 provides the detailed formulas for these statistics from a sample of results.

3.3 Comparative metrics calculated with the sampling approach

We now examine the case in which two or more samples exist, so in which a comparison of two or more products is at stake. The samples for a specific variable g_k can be written as $\{g_{k,1}\}$ and $\{g_{k,2}\}$; each of them is a series of values of length N. These samples can be used to calculate additional statistics and to test a number of hypotheses.

For the resulting sample $\{g_{k,1}\}$ and $\{g_{k,2}\}$, e.g., the carbon footprint of options 1 and 2, we can calculate two combinations of interest, the difference

$$\{d_k\} = \{g_{k,1,1} - g_{k,2,1}, g_{k,1,2} - g_{k,2,2}, \cdots, g_{k,1,N} - g_{k,2,N}\}. \tag{35}$$

and the ratio

$$\{r_k\} = \{g_{k,1,1}/g_{k,2,1}, g_{k,1,2}/g_{k,2,2}, \cdots, g_{k,1,N}/g_{k,2,N}\}. \tag{36}$$

Table 1 Key statistics that can be calculated from a series of sampling results A

Information	Parametric statistics	Non-parametric statistics
Location	Mean, $m(A)$	Median, $Q_2(A)$
Dispersion	Standard deviation, sd(A)	Interquartile range, IQR(A)
Relative dispersion	Coefficient of variation, CV(A)	Coefficient of quartile variation, CQV(A)
Range	95 % confidence interval, CI(A)	Range, $R(A)$
Distribution	Agreement with specified distribution (Kolmogorov-Smirnov statistic with z test)	–

The two samples $\{d_k\}$ and $\{r_k\}$ can be used to calculate the same statistics as listed in Table 1. In addition, tests of equality and correlation can be carried out; see Table 2 for some main statistics of interest.

3.4 Monte Carlo and LCA

The theory of Monte Carlo analysis is easily applied to LCA. In each run i, we draw all parameters from the specified distribution $(\mathbf{B};\mathbf{A};\mathbf{f})_i$ and calculate quantities of interest, such as $(\mathbf{g};\mathbf{s})_i$. We then subject these quantities to statistical analysis, by means of parametric or non-parametric statistics, using them for independent or dependent samples.

There are, however, certain issues to address. First, Monte Carlo is based on the repeated calculation of results. But if one calculation takes 1 min, repeating the calculation 10,000 times becomes time-consuming. Speed of calculation is therefore an important criterion to address.

Second, the number of Monte Carlo runs requires consideration. Morgan and Henrion (p. 200 ff.) discuss this issue. We can use the sample of results to study the convergence. For instance, we can study how the mean value m, the standard deviation (sd), or the coefficient of variation (CV) develops as a function of the sample size N:

$$m(z;N)=\frac{1}{N}\sum_{i=1}^{N} z_i;\ \mathrm{sd}(z;N)=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(z_i-m(z;N))^2};\ \mathrm{CV}(z;N)=\frac{\mathrm{sd}(z;N)}{m(z;N)} \quad (37)$$

Monitoring their convergence as N increases provides a clue to the sufficiency of the number of runs.

3.5 Contribution to variance

For the contribution to variance, the analytical method could employ explicit formulas for adding components to an overall variance. For the sampling methods, this will not work. Here, approaches that correlate the sample of input values with the sample of output values are available.

Another approach to calculating CTV is the one proposed by Geisler et al. (2005). Here, the correlation between each input parameter and the output parameter is calculated and used to partition the output uncertainty:

$$\mathrm{CTV}(z,x)=\frac{r^2(z,x)}{r^2(z,x)+r^2(z,y)}, \quad (38)$$

where in the case of Geisler et al. (2005) r refers to the rank (Spearman) correlation coefficient. Alternative approaches (Saltelli et al. 2001) are based on the Pearson correlation coefficient.

4 Three case studies

This section shows the results for three case studies. Section 4.1 is a small illustrative example, of which all details can be described in detail within this article. Section 4.2 is a large and realistic process LCA example, using the ecoinvent

Table 2 Statistics that can be calculated from two series of sampling results A and B. The null hypothesis is always that the values of the two samples are equal and the alternative hypothesis that they are unequal. The p statistic is therefore always based on a two-sided test

Information	Parametric statistics		Non-parametric statistics	
	Independent	Dependent	Independent	Dependent
Centrality	$m(A)=m(B)$ (t test)	$m(A)=m(B)$ (paired t test)	$Q_2(A)=Q_2(B)$ (Mann-Whitney with z test)	$Q_2(A)=Q_2(B)$ (Wilcoxon paired signed-rank with z test)
Correlation	–	$r(A,B)=0$ (Pearson correlation with t test)	–	$r_s(A,B)=0$ (Spearman correlation with t test)

v2 data with their uncertainties. Section 4.3 is a simple but large hybrid IO LCI example. Given that process analysis featured in the first two examples, emphasis is placed here on the IO part, and only minimal process information is included.

4.1 Illustrative example

For the first case, we refer to the example by Heijungs and Suh (2002, p. 14). This is an inventory with two processes, two products, and three elementary flows. The matrices are given by

$$\mathbf{A} = \begin{pmatrix} -2 & 100 \\ 10 & 0 \end{pmatrix}; \mathbf{B} = \begin{pmatrix} 1 & 10 \\ 0.1 & 2 \\ 0 & -50 \end{pmatrix}; \mathbf{f} = \begin{pmatrix} 0 \\ 1,000 \end{pmatrix} \quad (39)$$

All process inputs and emissions were given a normal distribution with a coefficient of variation of 10 % (Table 3).

In order to test the importance of the assumption of normality or log-normality, we repeated the calculation, now taking log-normal distributions with a geometric standard deviation of 1.3 (Table 4).

All computations can be done very quickly. Even the 100,000 Monte Carlo runs were finished in just a few seconds. As we see, both for normally distributed and for log-normally distributed data, the computations for analytical and sampling coincide very well. Our choice of presentation makes it once more clear that the Monte Carlo sampling approach can access more statistics (mean, quartiles, etc.) than the Taylor series approach, which can only calculate the standard deviation (and the variance of course).

We also used both methods to calculate the contribution to variance. Table 5 shows the results for the normally distributed uncertainties, using the analytical methods and the Monte Carlo method with 100,000 runs, both with a rank (Spearman) correlation (Geisler et al. 2005) and a normal (Pearson) correlation (Saltelli et al. 2001).

Again, the results are very similar. Each of the three methods assigns approximately equal contributions to the different input uncertainties.

Table 3 Statistics for the uncertainty of a very small LCA system with normally distributed uncertainties, computed by 100,000 Monte Carlo (MC) runs and by the analytical method (Taylor)

Flow	Baseline	m (MC)	sd (MC)	Q_2 (MC)	IQR (MC)	sd (Taylor)
CO_2	120	120	10.4	119	18.2	10.4
SO_2	14	14	1.15	13.9	2.02	1.15
Crude oil	−100	−100	14.1	−100	24.7	14.1

Table 4 Statistics for the uncertainty of a very small LCA system with log-normal distributed uncertainties, computed by 100,000 Monte Carlo (MC) runs and by the analytical method (Taylor)

Flow	Baseline	m (MC)	sd (MC)	Q_2 (MC)	IQR (MC)	sd (Taylor)
CO_2	120	120	13.7	119	18.2	13.7
SO_2	14	14	1.51	13.9	2.02	1.51
Crude oil	−100	−100	18.7	−100	24.7	18.6

4.2 Large process analysis using the ecoinvent database

The second case employs the ecoinvent database. The matrix **A** has size of 4,087×4,087, while **B** is 3,795×4,087. The ecoinvent consortium has moreover added impact categories to their data. Altogether, characterization factors for 672 impact categories are organized in a characterization matrix of size 672×3,795. We defined three alternatives, with reference flows "1 pkm transport, passenger car, RER," "1 pkm transport, aircraft, passenger, RER," and "1 pkm transport, high speed train, DE" and concentrated on the impact "kg CO_2-Eq IPCC 2007, climate change, GWP 100a, GLO." The data in ecoinvent comes with estimated uncertainties for almost all parameters. A total of 92,284 uncertainty distributions on **A** and **B** (out of 135,892, so 2 out of 3) were used, most log-normal. In addition, mock uncertainties were introduced on the characterization factors: normal with a CV of 10 %. The results for the three transport alternatives are in Table 6.

Again, the results of the analytical and the Monte Carlo approach are in good agreement with one another, in particular for the aircraft and train alternatives. For the car, the difference is larger, but we still judge the order of magnitude very comparable.

The analytical calculation finishes in a few minutes. This is in stark contrast with the Monte Carlo, which requires between 10 s and 1 min per run, depending on hardware and algorithm. Running 1,000 simulations thus takes between 2 and 16 h. Although 1,000 runs might be argued to be on the small side, the results are pretty stable and moreover in agreement with those of the analytical method. We did not try to go beyond the 1,000 runs.

We also did not complete a comparison of the contribution to variance by the two methods. With the analytical methods, the results were obtained in a few minutes. After 1,000 runs, the agreement with the Monte Carlo approach was very poor, and in fact, the CTVs showed to be very unstable, changing substantially from experiment to experiment. The correlations between the inputs and the output appear to require a much larger sample size than 1,000 to give a robust contribution to variance estimate. So, even though the uncertainty of the

Table 5 Contribution to variance (CTV) on the basis of 100,000 Monte Carlo (MC) runs, using the Pearson correlation coefficient and the rank (Spearman) correlation coefficient and using the analytical method (Taylor)

Process	Flow	CTV (MC, Pearson) (%)	CTV (MC, Spearman) (%)	CTV (Taylor) (%)
Electricity production	Fuel	4	4	4
Electricity production	CO_2	93	93	92
Fuel production	CO_2	3	4	4

result itself seems to be reliable with 1,000 runs, a contribution to variance seems to require many more runs.

4.3 Example using a hybrid IO-based LCI

The third case study is more extensive than the previous ones. It is different in several respects: it is IO-based instead of process-based, the functional unit is a consumption basket specified as the final demand block of the hybrid IO system, and an analysis of the effect of a first-order approximation is included.

4.3.1 Description of the system

In this example, we set up a closed input–output system

$$\begin{array}{|c|c|}\hline \mathbf{T} & \mathbf{y} \\ \hline \mathbf{v} & \mathbf{0} \\ \hline \end{array} \quad \begin{array}{|c|}\hline \mathbf{0} \\ \hline 1 \\ \hline \end{array},$$

$$\begin{array}{|c|c|}\hline \mathbf{B} & \mathbf{0} \\ \hline \end{array}$$

where **T** is a regional industry-by-industry input–output transaction table for Australia, and the vectors **v** and **y** describe the income earned and the products bought, respectively, by a hypothetical but typical Sydney household. **v** and **y** represent a single process integrated with the IO system. In input–output parlance, this system is termed "closed" because the usually exogenous income and expenditure vectors are endogenized into the intermediate transaction matrix. The alignment to the nomenclature introduced in Section 2 is via the intervention matrix **B** and the function unit **f**, as follows: The vector $\mathbf{f}^*=[\mathbf{0}|1]'$ selects and extracts one unit of one process (family metabolism) as the functional unit from the compound IO transaction matrix $\mathbf{T}^* = \begin{bmatrix}\mathbf{T} & \mathbf{y} \\ \mathbf{v} & \mathbf{0}\end{bmatrix}$. The matrix **B** holds the environmental interventions, as usual. **T** is sized 2,752×2,752 sectors (data sources are described in Gallego and Lenzen 2009). As a consequence, the extended matrices $\mathbf{T}^* = \begin{bmatrix}\mathbf{T} & \mathbf{y} \\ \mathbf{v} & \mathbf{0}\end{bmatrix}$, $\mathbf{A}^* = \mathbf{T}^*\widehat{\mathbf{x}*}^{-1}$, and $\mathbf{L}^*=(\mathbf{I}-\mathbf{A}^*)^{-1}$ are sized 2,753×2,753 sectors. The extended total use vector is $\mathbf{x}* = \begin{bmatrix}\mathbf{x} \\ z\end{bmatrix}$, with $z=\mathbf{y1}=\mathbf{1v}$ being the budget of the household and $\mathbf{x}=\mathbf{T1}+\mathbf{y1}$ being gross output of the economy. The satellite account **B** (content and data sources listed in ISA 2010) includes a row of total greenhouse gas emissions by industry, which we will employ in this case study to calculate the environmental intensity vector $\mathbf{b}^t$ (see Section 2.2.2). We ignore direct household emissions and set $\mathbf{b}^{t*}=[\mathbf{b}^t|0]$. **y** holds the expenditure on products by the average Sydney family, with a carbon footprint of about g_0 $z=52$ tonnes of CO_2-equivalents (data sources and Global Warming Potentials described in Lenzen and Peters 2010).

In our Monte Carlo experiments, we follow Hong et al. (2010) in specifying standard deviations $\Delta[\log_{10}(\mathbf{T})]$ in the logs of the input–output transactions matrix **T** and applying log-normal perturbations, i.e., $\widetilde{\mathbf{T}} = 10^{\log_{10}(\mathbf{T})\pm k\ \Delta[\log 10(\mathbf{T})]}$ with $k\in[0,1]$, so that the elements of $\widetilde{\mathbf{T}}$ can never become negative. For the sake of simplicity in this example, we assume that the environmental interventions $\mathbf{b}^t$ and the "process" vectors **v** and **y** and therefore also household budget z are known exactly. We also assume that total use **x** is certain (**x** is usually very large and therefore associated with much smaller uncertainties than the elements of **T**; cp. Bullard and Sebald 1977; 1988) so that the Monte Carlo perturbation only involves $\widetilde{\mathbf{A}^*} = \begin{bmatrix}\widetilde{\mathbf{T}} & \mathbf{y} \\ \mathbf{v} & \mathbf{0}\end{bmatrix}\widehat{\mathbf{X}^*}^{-1}$, $\widetilde{\mathbf{L}^*} = \left(\mathbf{I}-\widetilde{\mathbf{A}^*}\right)^{-1}$ and $\widetilde{g} = \mathbf{b}^{t*}\widetilde{\mathbf{L}^*}\mathbf{f}^*$.

$\widetilde{\mathbf{A}^*}$ is a compound matrix consisting of distinct sub-matrices. Since IO theory is usually formulated in terms of these sub-matrices, we will transform $\widetilde{g} = \mathbf{b}^{t*}\widetilde{\mathbf{L}^*}\mathbf{f}^*$ so that $\widetilde{g}$ is a separate function of $\widetilde{\mathbf{T}}$ and **y**. Utilizing Miyazawa's partitioned inverse (Miyazawa 1966), we find that

$$\left(\mathbf{I}-\widetilde{\mathbf{A}^*}\right)^{-1} = \begin{bmatrix}\mathbf{I}-\widetilde{\mathbf{T}}\widehat{\mathbf{x}}^{-1} & -\mathbf{y}z^{-1} \\ -\mathbf{v}\widehat{\mathbf{x}}^{-1} & \mathbf{I}\end{bmatrix}^{-1} = \begin{bmatrix}\widetilde{\mathbf{L}}\left(\mathbf{I}+\mathbf{y}z^{-1}\mathbf{K}\mathbf{v}\widehat{\mathbf{x}}^{-1}\widetilde{\mathbf{L}}\right) & \widetilde{\mathbf{L}}\mathbf{y}z^{-1}\mathbf{K} \\ \mathbf{K}\mathbf{v}\widehat{\mathbf{x}}^{-1}\widetilde{\mathbf{L}} & \mathbf{K}\end{bmatrix} \quad (40)$$

Table 6 Statistics for the uncertainty of a very large LCA system, computed by the analytical method and by 100 Monte Carlo runs

Alternative	Baseline	*m* (MC)	sd (MC)	Q_2 (MC)	IQR (MC)	sd (Taylor)
Car	0.182	0.181	0.0498	0.0173	0.0538	0.0557
Aircraft	0.126	0.126	0.00563	0.126	0.00862	0.0056
Train	0.0637	0.0649	0.00693	0.0652	0.0104	0.00657

with $\widetilde{\mathbf{L}} = \left(\mathbf{I}-\widetilde{\mathbf{T}}\widehat{\mathbf{x}}^{-1}\right)^{-1}$ and $\mathbf{K} = \left(\mathbf{I}-\mathbf{v}\widehat{\mathbf{x}}^{-1}\widetilde{\mathbf{L}}\mathbf{y}z^{-1}\right)^{-1}$. Considering that the household's budget is much smaller than the output of the economy, and therefore $\|\mathbf{v}\| \ll \|\mathbf{x}\|$, we find that $\mathbf{K} \approx \mathbf{I}$, and

$$\left(\mathbf{I}-\widetilde{\mathbf{A}}^{*}\right)^{-1} \approx \begin{bmatrix} \widetilde{\mathbf{L}} & \widetilde{\mathbf{L}}\mathbf{y}z^{-1} \\ \widehat{\mathbf{x}}^{-1}\widetilde{\mathbf{L}} & \mathbf{I} \end{bmatrix}. \tag{41}$$

With $\mathbf{f}^{*}=[\mathbf{0}|1]'$ and $\mathbf{e}^{*}=[\mathbf{e}|0]$, we arrive at $\widetilde{g} \approx \mathbf{b}^{t}\widetilde{\mathbf{L}}\mathbf{y}z^{-1}$, where term $\mathbf{y}z^{-1}$ contains the expenditure shares y_i/z in the total household budget z, nicely normalized to 1 as set in the functional unit.

4.3.2 Departure of Taylor approximation from the true inverse

First, we illustrate the departure of the first-order Taylor approximation (see Section 2.2.2) from the exact inverse formula, by plotting both the variations of the inverse-based carbon footprint

$$\Delta g^{\mathrm{Inv}} = \frac{\widetilde{g}\left(\widetilde{\mathbf{T}}\right)-g_0}{g_0}(\text{labeled`Inverse}') \tag{42}$$

and the variations of the carbon footprint based on a Taylor series approximation

$$\Delta g^{\mathrm{Taylor}} = \frac{\sum_{ij}\left(\widetilde{T}_{ij}-T_{0,ij}\right)\dfrac{\left[\mathbf{b}^{t}\left(\mathbf{I}-\mathbf{T}_0\widehat{\mathbf{x}}^{-1}\right)^{-1}\right]_i\left[\left(\mathbf{I}-\mathbf{T}_0\widehat{\mathbf{x}}^{-1}\right)^{-1}\mathbf{y}z^{-1}\right]_j}{x_j}}{g_0} \quad (\text{labeled 'Taylor'}) \tag{43}$$

into the same diagram (Fig. 2).

Up to about $k=\pm\frac{1}{2}$, or $\widetilde{\mathbf{T}} = 10^{\log_{10}(\mathbf{T})\pm\frac{1}{2}\Delta[\log_{10}(\mathbf{T})]}$, the Taylor approximation gives reasonable results; however, beyond $|k|>\frac{1}{2}$, deviations become significant to the extent that the uncertainty in g is underestimated for $k>0$ and overestimated for $k<0$. This result can be explained by the fact that the Taylor approximation neglects terms of second and higher order, and hence, the curvature of the Taylor approximation is smaller.

In the calculation above, all elements of $\widetilde{\mathbf{T}}$ departed from those of $\mathbf{T}$ in one and the same direction. In reality, the "true" values of some elements in $\mathbf{T}$ may be larger than the nominal values, and others may be smaller, so that effects on the overall uncertainty of g will cancel out. This behavior can either be simulated using Monte Carlo analysis or analytically approximated using the error propagation formula given in Section 2.2.2.

4.3.3 Analytical approach to uncertainty

The analytical approximation proceeds as follows: We start with the variance formula from Section 2.2, ignoring covariance terms:

$$\mathrm{var}[g] \approx \sum_{ij}\left(\frac{dg}{dT_{ij}}\right)^{2}\mathrm{var}\left[\log_{10}\left(T_{ij}\right)\right]. \tag{44}$$

Starting with our earlier approximation $g \approx \mathbf{b}^{t}\widetilde{\mathbf{L}}\mathbf{y}z^{-1}$, we express g in terms of $\log_{10}(\mathbf{T})$ as

$$g(\log_{10}(\mathbf{T})) \approx \mathbf{b}^{t}\left(\mathbf{I}-10^{\log_{10}(\mathbf{T})}\widehat{\mathbf{x}}^{-1}\right)^{-1}\mathbf{y}z^{-1}. \tag{45}$$

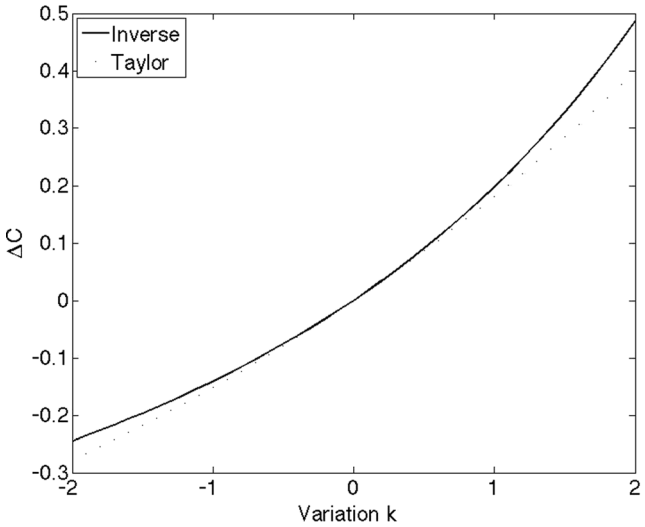

Fig. 2 Δg^{Inv} (*solid line*) and $\Delta g^{\mathrm{Taylor}}$ (*dotted line*) as a function of the variation parameter k in $\widetilde{\mathbf{T}} = 10^{\log_{10}(\mathbf{T}) \pm k\ \Delta[\log 10(\mathbf{T})]}$

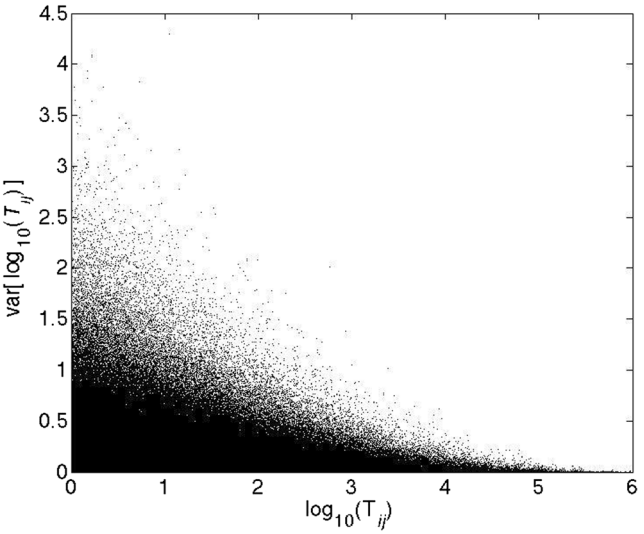

Fig. 3 Data for var[$\log_{10}(T_{ij})$] plotted against $\log_{10}(T_{ij})$ (data taken from Gallego and Lenzen 2009; Lenzen et al. 2010)

The derivative in the variance formula can be worked out by applying the chain rule:

$$\frac{\partial \mathrm{g}(T_{ij})}{\partial \log_{10}(T_{ij})} = \frac{\partial \mathrm{g}(T_{ij})}{\partial T_{ij}} \frac{\partial\left(10^{\log_{10}(T_{ij})}\right)}{\partial \log_{10}(T_{ij})} = \frac{\partial \mathrm{g}(T_{ij})}{\partial T_{ij}} \frac{\partial\left(\mathrm{e}^{\ln(10)\ \log_{10}(T_{ij})}\right)}{\partial \log_{10}(T_{ij})}$$
$$= \frac{\left[\mathbf{b}^{\mathrm{t}}\left(\mathbf{I}-\mathbf{T}\widehat{\mathbf{x}}^{-1}\right)^{-1}\right]_i \left[\left(\mathbf{I}-\mathbf{T}\widehat{\mathbf{x}}^{-1}\right)^{-1}\mathbf{y}\right]_j}{x_j} \ln(10) T_{ij} \tag{46}$$

Inserting into the variance formula yields

$$\mathrm{var}[g] \approx \sum_{ij}\left(\frac{\left[\mathbf{b}^{\mathrm{t}}(\mathbf{I}-\mathbf{T}\widehat{\mathbf{x}}^{-1})^{-1}\right]_i \left[(\mathbf{I}-\mathbf{T}\widehat{\mathbf{x}}^{-1})^{-1}\mathbf{y}\right]_j}{x_j} \ln(10) T_{ij}\right)^2 \mathrm{var}\left[\log_{10}(T_{ij})\right]$$
$$= \sum_{ij}\left[(\mathbf{b}^{\mathrm{t}}\mathbf{L})_i (\mathbf{L}\mathbf{y})_j \ln(10) A_{ij}\right]^2 \mathrm{var}\left[\log_{10}(T_{ij})\right]. \tag{47}$$

Standard deviations and coefficients of variation can be calculated from the above expression for the variance according to formulae derived above. The input data for the term $\mathrm{var}\left[\log_{10}(T_{ij})\right]$ (Fig. 3) were computed by applying a RAS-type approach to the variance of primary data used to construct the IO system (for further details on the RAS approach, see Gallego and Lenzen 2009). In this RAS approach, an error propagation formula is fitted to data for the standard deviations of primary data, as a function of the standard deviations of the IO table elements T_{ij}. Figure 3 shows that variances are large for small elements T_{ij} and vice versa (compare with Lenzen et al. 2010).

Inserting all data, the Taylor approximation yields $\mathrm{CV}(g) = \frac{\mathrm{sd}(g)}{g} = \frac{\sqrt{\mathrm{var}[g]}}{g} = 2.18\ \%$. The calculation for the 2,752×2,752-sized system takes about 20 s on an off-the-shelf laptop equipped with a 1.6-GHz Intel Core.

4.3.4 Sampling approach to uncertainty

The sampling approach proceeds as follows: Applying Monte Carlo simulation requires the repeated calculation of $\widetilde{\mathbf{T}} = 10^{\log_{10}(\mathbf{T}) + k\sqrt{\mathrm{var}[\log_{10}(\mathbf{T})]}}$, where $k \in [-1,1]$ is a normally distributed random variable, calculating $\widetilde{g} \approx \mathbf{b}^{\mathrm{t}}\left(\mathbf{I}-\widetilde{\mathbf{T}}\widehat{\mathbf{x}}^{-1}\right)^{-1}\mathbf{y}$ z^{-1} from $\widetilde{\mathbf{T}}$, plotting the distribution of $\frac{\mathrm{sd}(g)}{g} = \frac{\sqrt{\mathrm{var}[g]}}{g} = \frac{\widetilde{g}-g_0}{g_0}$, and finally deriving var[g] from the mean of this distribution.

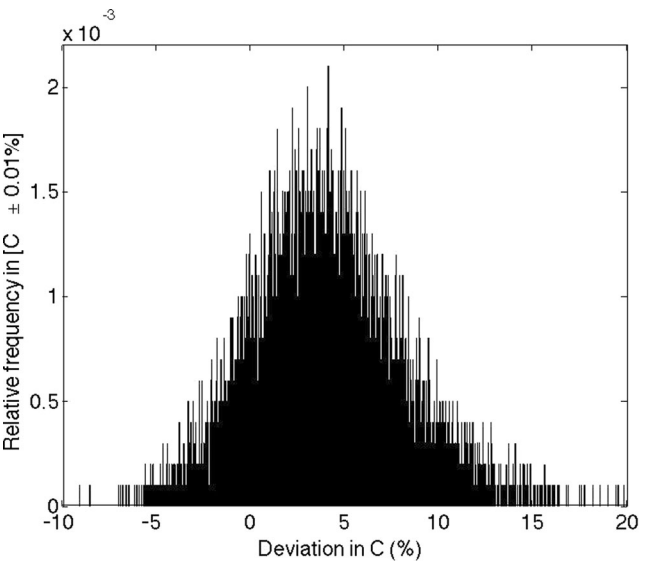

Fig. 4 Frequency distribution of relative deviations $\frac{\widetilde{g}-g_0}{g_0}$. Frequency intervals ("bins") are 0.01 % apart, and the distribution peak occurs at $\frac{\mathrm{sd}(g)}{g} = 2.06\ \%$

Table 7 Summary of statistics that can be calculated from the results of a sampling approach

Sample	Parametric statistics	Non-parametric statistics
One sample or independent samples	**Mean** (m)	Median (Q_2)
	Standard deviation (sd)	Interquartile range (*IQR*)
	Coefficient of variation (CV)	Coefficient of quartile variation (*CTV*)
	Confidence intervals (CI)	Range (R)
	p (mean=0)	$p(Q_2=0)$
Dependent samples	**Mean difference** (m_d)	Median difference (Q_{2d})
	Mean ratio (m_r)	Median ratio (Q_{2r})
	$p(m_d=0)$	$p(Q_{2d}=0)$

The statistics that are also available with the analytical method are presented in bold

The result of 100,000 Monte Carlo runs on this system (Fig. 4) yield a mean for $\mathrm{CV}(g) = \frac{\mathrm{sd}(g)}{g} = \frac{\sqrt{\mathrm{var}[g]}}{g} = 2.06\ \%$. One perturbation run of the 2,752×2,752-sized system takes just over 5 s on an off-the-shelf laptop equipped with a 1.6-GHz Intel Core, about the same time per element (0.6 μs) as for the slightly larger ecoinvent example when running for 10 s (Section 4.2). One hundred thousand runs of this system took just over 6 days. This shows that even parallel execution of multiple cores will not significantly change the odds of Monte Carlo codes in terms of computer runtime.

These results support our conclusion that an analytical approach will yield results that are very close (within about 5 % in this case) to the sampling-based results, but that can be obtained in a vastly shorter time.

5 Conclusions and discussion

In this paper, we focused on a comparison of two classes of error propagation methods, the analytical and the sampling approach. We discussed their foundation and implementation in LCA, in terms of the data input needs, the formulas, and the types of output obtained, with an emphasis on the differences and similarities in performance.

Our first conclusion is that the two approaches differ in terms of their required input data. The sampling procedure requires a specification of the probability distributions of the input data, for instance in terms of a normal or log-normal distribution. Such a specification requires a modeller or database provider to estimate such distributions for each parameter. Traditionally, users specify a central value or only a mean or median value, which is then typically interpreted as the most likely value. In specifying a probability distribution, a user needs to provide at least two more types of information: the shape of the distribution and the parameter or parameters describing that distribution. For instance, a user could specify the shape as normally distributed and a parameter for the standard deviation, the variance, or the SD95 (range within which 95 % of the points are located). In cases of the most widely used distributions (uniform, symmetric triangular, log-normal), one parameter for dispersion suffices. In the case of other distributions (non-symmetric triangular, beta, Weibull, etc.), additional parameters are needed.

For the analytical approach, only the second moment of the distribution (i.e., the variance) is needed. This means that the distribution itself need not be known, and even if it were known that it would not need to be passed to the software. Thus, the data requirements are smaller for the analytical than for the sampling approaches.

A second conclusion is that the propagation algorithms are the most distinguished features of the two approaches. The default approach for the analytical method, studied in this paper, is the Gaussian method, based on a first-order Taylor series approximation. The default approach for the sampling method is the Monte Carlo method. Both methods are easy to implement in software.

A third conclusion is that the two methods provide different sets of information. The analytical method does not calculate a distribution but only yields the second moment (the variance) of the distribution. Of course, this gives an important clue to the confidence intervals, as it is the most basic indicator of dispersion.

The sampling method returns a sample of results, from which many statistics can be calculated. In this paper, we discussed two-by-two classes of such statistics: parametric vs. non-parametric and independent vs. dependent samples (see Table 7).

Next, in a numerical comparison, the two methods give similar results, at least to the extent that they yield shared indicators. Basically, the only shared indicator between the analytical and sampling approach is the variance (and thereby the standard deviation and coefficient of variation). Many of the statistics that are accessible by the sampling approach (median, interquartile range, etc.) cannot be obtained using the analytical approach. That is an important aspect for a user facing the choice between an analytical and a sampling approach. Another distinguishing feature is that the analytical approach is based on a linearization that only holds when the error term is not too big, while the sampling approach may perform better, albeit with a larger number of runs. Covariances between random input variables can in principle be included in both approaches, but we conjecture that, except

for small and tailor-made systems, such covariances will be unknown, and their propagation will be unfeasible in practice.

On the other hand, there is an enormous performance difference in terms of computing time. For small systems (with a few unit processes), this difference is unimportant, but for larger systems (of the size of ecoinvent), it can mean days instead of minutes. More specifically, variances calculated using sampling and analytical approaches for our small system agree with one another, but for the large system, the sampling method would take so much time that an analytical approach seems most appropriate.

The take-home message is subtle. Sampling methods give access to more types of information (because an entire distribution is calculated instead of just two moments of the distribution) than analytical methods do, but they require more information and (sometimes much) more computer time. Most, but not all, key information can more rapidly be extracted using an analytical method. Tests of significant differences between life cycle scenarios, one of the most critical results obtained from LCA studies, have so far only been undertaken with sampling methods. Our work shows that the comparative advantages of analytical methods would enable such tests to be carried out at much reduced expense, ultimately leading to enhanced life cycle information. At the same time, it should be stressed that uncertainty analyses have intrinsic limitations. They should be handled with care, and they must be supplemented by sensitivity analyses.

Acknowledgments This article has been written in the context of the UNEP/SETAC working group on uncertainty. Members of this group include, besides the authors of this article, Andreas Ciroth, Ralph Rosenbaum, Mark Huijbregts, Fausto Freire, Tom McKone, Olivier Jolliet, and Enrico Benetto. The group's activities and the writing of this article have been made possible through the financial support of The Sustainability Consortium. This work was also financially supported by the National eResearch Collaboration Tools and Resources project (NeCTAR) through its Industrial Ecology Virtual Laboratory and by the Australian Research Council through its Discovery Projects DP0985522 and DP130101293. Two anonymous referees have provided valuable comments that helped improving on an early draft of this article.

Appendix A: Formulas for sampling statistics

Table 8 The sample of size N is indicated as $\{A_i\}_{i=1,\ldots,N}=\{A_1,\ldots,A_N\}$. The ordered sample is indicated by $\left\{\widetilde{A}_i\right\}_{i=1,\ldots N}$.

Statistic	Symbol	Formula
Mean	m	$m(A)=\frac{1}{N}\sum_{i=1}^{N} A_i$
Variance	var	$\mathrm{var}(A)=\frac{1}{N-1}\sum_{i=1}^{N}(A_i-m)^2$
Standard deviation	sd	$\mathrm{sd}(A)S=\sqrt{\mathrm{var}(A)}$
Coefficient of variation	CV	$\mathrm{CV}(A)=\frac{\mathrm{sd}(A)}{m(A)}$
95 % confidence interval	CI	$\mathrm{CI}(A)=[m(A)-1.96\mathrm{sd}(A),m(A)+1.96\mathrm{sd}(A)]$
Median	Q_2	$Q_2(A)=\begin{cases}\widetilde{A}_{(N+1)/2} & N\ odd\\ \left(\widetilde{A}_{N/2}+\widetilde{A}_{N/2+1}\right)/2 & N\ even\end{cases}$
First quartile	Q_1	$Q_1(A)=\begin{cases}Q2\left\{\widetilde{A}_i\right\}_{i=1,\ldots,(N+1)/2} & N\ \mathrm{odd}\\ Q2\left\{\widetilde{A}_i\right\}_{i=1,\ldots,N/2} & N\ \mathrm{even}\end{cases}$
Third quartile	Q_3	$Q_3(A)=\begin{cases}Q2\left\{\widetilde{A}_i\right\}_{i=(N+1)/2,\ldots,N} & N\ \mathrm{odd}\\ Q2\left\{\widetilde{A}_i\right\}_{i=N/2,\ldots,N} & N\ \mathrm{even}\end{cases}$
Interquartile range	IQR	$\mathrm{IQR}(A)=Q_3(A)-Q_1(A)$
Coefficient of quartile variation	CQV	$\mathrm{CQV}(A)=\frac{\mathrm{IQR}(A)}{Q_3(A)+Q_1(A)}$
Range	R	$R(A)=\widetilde{A}_N-\widetilde{A}_1$
Mean difference	m_d	$m_d(A,B)=m(A)-m(B)$
Mean ratio	m_r	$m_r(A,B)=\frac{1}{N}\sum_{i=1}^{N} A_i/B_i$
Correlation	r	$r(A,B)=\frac{\sum_{i=1}^{N}(A_i-m(A))(B_i-m(B))}{\sqrt{\sum_{i=1}^{N}(A_i-m(A))^2}\sqrt{\sum_{i=1}^{N}(B_i-m(B))^2}}$
Rank correlation	r_s	$r_s(A,B)=\frac{\sum_{i=1}^{N}\left(\widetilde{A}_i-Q_2(A)\right)\left(\widetilde{B}_i-Q_2(B)\right)}{\sqrt{\sum_{i=1}^{N}\left(\widetilde{A}_i-Q_2(A)\right)^2}\sqrt{\sum_{i=1}^{N}\left(\widetilde{B}_i-Q_2(B)\right)^2}}$

References

Apostol TM (1967) Calculus, Vol. 1. One-variable calculus, with an introduction to linear algebra. Wiley

Ayyub BM, Klir GJ (2006) Uncertainty modeling and analysis in engineering and the sciences. Chapman & Hall/CRC, Boca Raton

Bevington PR, Robinson DK (1994) Data reduction and error analysis for the physical sciences. McGraw-Hill, Inc., New York

Bullard CW, Penner PS, Pilati DA (1978) Net energy analysis—handbook for combining process and input-output analysis. Resour Energ 1:267–313

Bullard CW, Sebald AV (1977) Effects of parametric uncertainty and technological change on input-output models. Rev Econ Stat LIX, pp 75-81

Bullard CW, Sebald AV (1988) Monte Carlo sensitivity analysis of input-output models. Rev Econ Stat LXX, pp 708-712

Ciroth A (2001) Fehlerrechnung in Ökobilanzen. PhD thesis, Technischen Universität Berlin, Berlin

Ciroth A, Fleischer G, Steinbach J (2004) Uncertainty calculation in life cycle assessments. A combined model of simulation and approximation. Int J Life Cycle Assess 9(4):216–226

Crama Y, Defourny J, Gazon J (1984) Structural decomposition of multipliers in input-output or social accounting matrix analysis. Econ Appl 37:215–222

Defourny J, Thorbecke E (1984) Structural path analysis and multiplier decomposition within a social accounting matrix framework. Econ J 94:111–136

Gallego B, Lenzen M (2009) Estimating generalised regional input-output systems: a case study of Australia. In: Ruth M, Davíðsdóttir B (eds) The dynamics of regions and networks in industrial ecosystems. Edward Elgar Publishing, Boston, pp 55–82

Geisler G, Hellweg S, Hungerbühler K (2005) Uncertainty analysis in life cycle assessment (LCA): case study on plant-protection products and implications for decision making. Int J Life Cycle Assess 10: 184–192

Heijungs R, Guinée JB, Huppes G, Lankreijer RM, Udo de Haes HA, Wegener Sleeswijk A, Ansems AMM, Eggels PG, van Duin R, de Goede HP (1992) Environmental life cycle assessment of products. I: Guide—October 1992. II: Backgrounds—October 1992. CML (NOH report 9266 + 9267; ISBN 90-5191-064-9), Leiden, 1992, 96 + 130 pp

Heijungs R (1994) A generic method for the identification of options for cleaner products. Ecol Econ 10(1):69–81

Heijungs R (1996) Identification of key issues for further investigation in improving the reliability of life-cycle assessments. J Clean Prod 4: 159–166

Heijungs R, Suh S (2002) The computational structure of life cycle assessment. Kluwer Academic Publishers, Dordrecht

Heijungs R, Huijbregts MAJ (2004) A review of approaches to treat uncertainty in LCA. p. 332-339. In: Pahl-Wostl C, Schmidt S, Rizzoli AE, Jakeman AJ (eds) Complexity and integrated resources \management. Transactions of the 2nd Biennial Meeting of the International Environmental Modelling and Software Society, Volume 1. iEMSs (ISBN 88-900787-1-5), Osnabrück. 2004, 1533 pp

Heijungs R, Suh S, Kleijn R (2005) Numerical approaches to life cycle interpretation. The case of the ecoinvent'96 database. Int J Life Cycle Assess 10:103–112

Heijungs R (2010) Sensitivity coefficients for matrix-based LCA. Int J Life Cycle Assess 15:511–520

Hong J, Shaked S, Rosenbaum R, Jolliet O (2010) Propagation of uncertainty in life cycle inventory and impact assessment: application to an automobile front panel. Int J Life Cycle Assess 15(5): 499–510

Huijbregts MAJ (1998) Application of uncertainty and variability in LCA. Part II: Dealing with parameter uncertainty and uncertainty due to choices in life cycle assessment. Int J Life Cycle Assess 3(6): 343–351

Ii R (2000) Changes of uncertainty in life-cycle inventory analysis with I-O tables of different years. The Fourth International Conference on EcoBalance. Tsukuba, Japan, 197-200

Imbeault-Tétrault H, Jolliet O, Deschênes L, Rosenbaum RK (2013) Analytical propagation of uncertainty in life cycle assessment using matrix formulation. J Ind Ecol 17:485–492

ISA (2010) *Indicators*. Information Sheet 6, Internet site http://www.isa.org.usyd.edu.au/research/InformationSheets/ISATBLInfo6_v1.pdf, Sydney, Australia, ISA, The University of Sydney

Lenzen M (2000) Errors in conventional and input-output-based life-cycle inventories. J Ind Ecol 4:127–148

Lenzen M (2002) A guide for compiling inventories in hybrid LCA: some Australian results. J Clean Prod 10:545–572

Lenzen M (2006) Uncertainty of end-point impact measures: implications for decision-making. Int J Life Cycle Assess 11:189–199

Lenzen M, Crawford RH (2009) The path exchange method for hybrid LCA. Environ Sci Technol 43:8251–8256

Lenzen M, Peters G (2010) How city dwellers affect their resource hinterland—a spatial impact study of Australian households. J Ind Ecol 14:73–90

Lenzen M, Wood R, Wiedmann T (2010) Uncertainty analysis for multi-region input-output models—a case study of the UK's carbon footprint. Econ Sys Res 22:43–63

Lloyd SM, Ries R (2007) Characterizing, propagating, and analyzing uncertainty in life-cycle assessment. J Ind Ecol 11(1):161–179

Miyazawa K (1966) Internal and external matrix multipliers in the input-output model. Hitotsub J Econ 7:38–55

Morgan MG, Henrion M (1990) Uncertainty. A guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press, Cambridge

Moskowitz PD, Rowe MD (1985) A comparison of input-output and process analysis. In: Ricci PF, Rowe MD (eds) Health and environmental risk assessment. NY, USA, Pergamon Press, New York, pp 281–293

Nansai K, Tohno S, Kasahara M (2001) Uncertainty of the embodied CO_2 emission intensity and reliability of life cycle inventory analysis by input-output approach. Energ Resour 22, (published by the Japan Society of Energy and Resources, in Japanese)

Peters G (2007) Efficient algorithms for life cycle assessment, input-output analysis, and Monte-Carlo analysis. Int J Life Cycle Assess 12(6):373–380

Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1992) Numerical recipes in Pascal. The art of scientific computing. Cambridge University Press, Cambridge

Quandt RE (1958) Probabilistic errors in the Leontief system. Naval Res Logistics Quarterly 5:155–170

Quandt RE (1959) On the solution of probabilistic Leontief systems. Nav Res Log 6:295–305

Sakai S, Tanno S, Hondo H (2000) Uncertainty analysis for I/O analysis using perturbation method. The Fourth International Conference on EcoBalance. Tsukuba, Japan, pp 189-192

Saltelli A, Chan K, Scott EM (2001) Sensitivity analysis. John Wiley & Sons, Chichester

Siegel S (1956) Nonparametric statistics for the behavioral sciences. McGraw-Hill Book Company, New York

Suh S (2004) Functions, commodities and environmental impacts in an ecological-economic model. Ecol Econ 48:451–467

Suh S, Heijungs R (2007) Power series expansion and structural analysis for life cycle assessment. Int J Life Cycle Assess 12: 381–390

Suh S, Huppes G (2005) Methods for life cycle inventory of a product. J Clean Prod 13:687–697

Suh S, Lenzen M, Treloar GT, Hondo H, Horvath A, Huppes G, Jolliet O, Klann U, Krewitt W, Moriguchi Y, Munksgaard J, Norris G (2004) System boundary selection in life-cycle inventories. Environ Sci Technol 38:657–664

Suh S, Nakamura S (2007) Five years in the area of input-output and hybrid LCA. Int J Life Cycle Assess 12:351–352

Treloar G (1997) Extracting embodied energy paths from input-output tables: towards an input-output-based hybrid energy analysis method. Econ Syst Res 9:375–391

Williams ED, Weber CL, Hawkins TR (2009) Hybrid approach to characterising and managing uncertainty in life cycle inventories. J Ind Ecol, in press

Yoshida Y, Ishitani H, Matsuhashi R, Kudoh Y, Okuma H, Morita K, Koike A, Kobayashi O (2002) Reliability of LCI considering the uncertainties of energy consumptions in input-output analyses. Appl Energ 73:71–82

Yoshida Y, Ishitani H, Matsuhashi R, Kudoh Y, Okuma H, Morita K, Koike A, Kobayashi O, Yoshioka K (2001) Evaluation of uncertainty in LCA based on input-output analysis. Technol 8